

Online leak diagnosis in pipelines using an EKF-based and steady-state mixed approach

I. Santos-Ruiz^a, J.R. Bermúdez^a, F.R. López-Estrada^{a,*}, V. Puig^b, L. Torres^{c,d}, J.A. Delgado-Aguíñaga^e

^a Tecnológico Nacional de México/Instituto Tecnológico de Tuxtla Gutiérrez, TURIX-Dynamics Diagnosis and Control Group., Carretera Panamericana km 1080 SN, Tuxtla Gutiérrez, México, CP 29050, Mexico

^b Department of Automatic Control (ESAII), Universitat Politècnica de Catalunya (UPC), Rambla de Sant Nebridi 10, 08222 Terrassa, Spain

^c Universidad Nacional Autónoma de México, Instituto de Ingeniería, Ciudad de México, Mexico

^d Cátedras CONACyT, Mexico

^e Universidad del Valle de México, Campus Guadalajara Sur, CIIDETEC, 45601 San Pedro Tlaquepaque, Jalisco, Mexico

ARTICLE INFO

Keywords:

Pipelines
Leak diagnosis
Extended Kalman Filter
Data fusion
Steady state
Real-time

ABSTRACT

This paper proposes a methodology for leak detection and isolation (LDI) in pipelines based on data fusion from two approaches: a steady-state estimation and an Extended Kalman Filter (EKF). The proposed method considers only pressure head and flow rate measurements at the pipeline ends, which contain intrinsic sensor and process noise. The LDI system is tested in real-time by using a USB data acquisition device that is implemented in MATLAB environment. The effectiveness of the method is analyzed by considering: online detection, location as well as quantification of non-concurrent leaks at different positions. The leak estimation error average is less than 1% of the flow rate and less than 3% in the leakage position. Furthermore, the incorporation of a steady-state estimation shows that the solution of the LDI problem has improved significantly with respect to the one that only considers the EKF estimation. An experimental analysis was also performed on the effectiveness of the proposed approach for different sampling rates and for different leakage positions.

1. Introduction

In fluid distribution systems, automatic fault monitoring and diagnosis are of great relevance worldwide. The primary purpose of an automatic pipeline monitoring system is to detect leaks, obstructions or sensor faults as quickly as possible, with a minimum of instrumentation and cost (Verde & Torres, 2017). In the case of leaks, these can cause substantial economic losses, damage to the environment and health risks. To size the problem, in water distribution networks the worldwide percentage of volumetric leakage losses has been estimated at around 21%, although in countries such as Mexico it reaches an average value of 40% (OECD, 2016). Also in Mexico, the state-owned oil company Pemex loses about two billion US dollars per year due to leaks in pipelines caused by clandestine outlets of the so-called “huachicoleros”, which also pollute the environment and have caused explosions with the loss of human lives (Hernández, 2017; Oswald, 2017).

There are several methods for the direct detection of leaks, which are based on visual or palpable physical detection of the fluid such as hardware-based or Computational Pipeline Monitoring (CPM) methods.

Hardware-based methods depend heavily on the physical equipment installed along the pipeline. On the other hand, CPM refers to software-based systems that operate with limited instrumentation and provide algorithmic tools that expand the possibilities of pipeline operators to recognize anomalies that can help to detect the leaks (API, 2002). The CPM tools are based on mathematical models of the pipeline and are complemented with measurement data of some physical variables associated with the flow process e.g. pressure, flow rate, and temperature, among others.

Although it is unavoidable that a leak detection technique depends on both, the mathematical models and the data processing, in the literature usually distinguish two approaches. On the one hand, data-driven detection methods that focus on the digital processing of signals prioritizing the statistical analysis of measurements and can be consulted in Mashford, Silva, Marney, and Burn (2009), Arifin, Li, Shah, Meyer, and Colin (2018), Camacho-Navarro, Ruiz, Perez, Villamizar, and Mujica (2015), Soldevila, Fernandez-Canti, Blesa, Tornil-Sin, and Puig (2017). On the other hand, model-based methods incorporate

* Corresponding author.

E-mail address: frlopez@ittg.edu.mx (F.R. López-Estrada).



Fig. 1. Experimental pipeline.

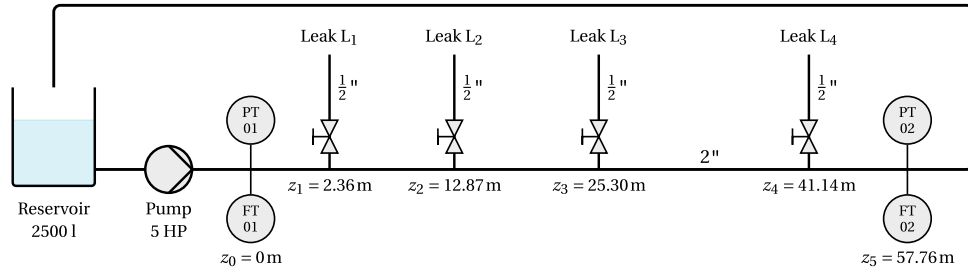


Fig. 2. P&I diagram of experimental pipeline.

dynamic equations based on physical principles (Delgado-Aguíñaga, Besançon, Begovich, & Carvajal, 2016; Torres, Verde, Besançon, & González, 2014; Verde, 2001, 2004). There are also proposals for mixed methods that combine intensive data processing and mathematical flow models (Soldevila et al., 2016). Recently, comprehensive taxonomies of leak detection systems were presented in Henrie, Carpenter, and Nicholas (2016), Murvay and Silea (2012). According to these classifications, our proposal is included within the model-based techniques.

In the leak diagnosis, not only it is important to detect the leak and quantify it, but it is a priority to locate it as accurately as possible. So, it is necessary to have algorithms that precisely determine the location of the leak because it is not always visible from outside the pipe. An error of a few meters in locating a leak in an underground pipeline results in a significant cost for digging in a wrong area to repair it (Delgado-Aguíñaga & Begovich, 2017). In practice, one circumstance that limits the precise location of leakage using approaches based on the deterministic dynamic models of the pipeline flow is the presence of variable parameters in the mathematical model, which show a nonlinear dependence on the flow rate and other physical factors such as temperature (Dulhoste, Guillén, Besançon, & Santos, 2017). Other complications occur because the signals obtained from the pressure and flow sensors are generally noisy (Billmann & Isermann, 1987).

This paper proposes a methodology for detection, location, and estimation of fluid leaks in a straight pipeline (without branchings) by considering pressure and flow measurements at the pipeline ends. This set-up is justified because many hydraulic and fuel transport networks are instrumented in this way (Verde, Gentil, & Morales-Mené, 2013). Considering the noisy characteristic of the available measurements, this work is based on the proposal of Delgado-Aguíñaga, Besançon, and Begovich (2015), when using an EKF estimator, but also it incorporates the steady-state solution of the pipeline dynamic model to refine the solution and it proposes a strategy to accelerate the convergence of the iterative process. These added features are not only desirable but necessary since real-time implementation requires fast and accurate LDI methods. A feature that is highlighted is the live implementation using conventional hardware and software, unlike most other works

referred in the literature that are limited to presenting simulation results or offline calculations. A deep analysis concerning the sensitivity of the system, its reliability and its accuracy in the location of the leaks are commented. An exhaustive experimental analysis was performed with 96 different leak scenarios, which illustrates the effectiveness and applicability of the proposed method.

Regarding the delimitation of the conditions where the proposal is applicable, only pipelines with pressurized flow are considered. Furthermore, the topic of LDI in biphasic flows has not been considered in this work, because the mathematical model used by the EKF would change due to phase transitions, so the current proposal does not include this possibility.

2. Materials and methods

2.1. Description and modeling of experimental pipeline

The experimental tests of the proposed method for leak detection and localization that is presented in this work were carried out on a pilot pipeline plant located at Laboratory of Hydraulics in the Tuxtla Gutiérrez Institute of Technology. This experimental pipeline is serpentine-shaped (Fig. 1) with an equivalent length of 57.76 m and it is made of 2-inch diameter PVC; the water is driven with a centrifugal pump whose power is controlled by a frequency inverter. A more detailed description of this pipeline setup is presented in Bermúdez, Santos-Ruiz, López-Estrada, Torres, and Puig (2017).

A P&I diagram of the pipeline setup used is shown in Fig. 2. Industrial pressure and flow-rate sensors/transmitters are located at the ends of the system. These sensors provide the information that will be used as inputs and outputs for the EKF estimator. Four manual gate valves in half-inch tees are located at arbitrary positions to simulate leakages, which are modeled as orifice outlets.

The signals from the pressure and flow transmitters are received through 4–20 mA current loops connected to a 14-bit data acquisition system (DAQ) with USB interface to MATLAB. As shown in Fig. 3, the current signals are read in voltage form through 470 Ω resistors at the

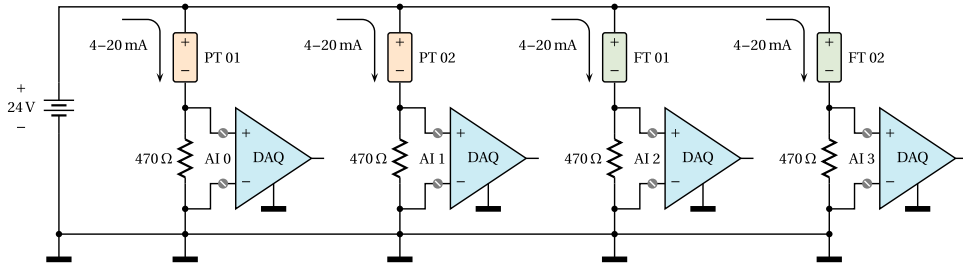


Fig. 3. Pressure/flow transmitters to DAQ electrical connections.

Table 1

Parameters of experimental pipeline.

Parameter	Value
Length, L	57.76 m
Diameter, d	0.052 m
Relative roughness, ϵ	2.47×10^{-4}
Fluid (water) density, ρ	995.736 kg/m ³
Kinematic viscosity, ν	8.03×10^{-7} m ² /s
Wave speed, c	422.754 m/s
Gravity acceleration, g	9.782 999 m/s ²

DAQ inputs. The wiring of the current loops is made with 20 AWG twisted pair. A MATLAB GUI (graphical user interface) continuously displays the pressure head and flow rate signals on the screen and it shows alarms with the details of leaks when they occur. The diagnosis software also includes data logging functions for the subsequent analysis of results offline.

The leak detection and localization method proposed in this paper is formulated from a dynamic model of the pipeline variables h (piezometric head, a quantity proportional to the pressure) and q (volumetric flow rate). A set of first-order partial differential equations consisting of the momentum Eq. (1) and the continuity Eq. (2) are derived in Chaudhry (2014) applying the momentum and mass conservation principles to a control volume considering a horizontal pipeline of constant diameter d and cross sectional area A :

$$\frac{\partial q(z, t)}{\partial t} + gA \frac{\partial h(z, t)}{\partial z} + \frac{f(z, t)}{2Ad} q(z, t) |q(z, t)| = 0, \quad (1)$$

$$\frac{\partial h(z, t)}{\partial t} + \frac{c^2}{gA} \frac{\partial q(z, t)}{\partial z} = 0, \quad (2)$$

where z is the position variable directed along the pipeline axis, from the supply (inlet) to the delivery (outlet) point, and t is the time variable defined from an arbitrary clock reference. The above equations are hyperbolic-type, so they describe the propagation of waves in the fluid. The other literals in these equations are the friction factor f , the wave velocity c , and the acceleration due to gravity g .

To compute the friction factor f , a prior value was determined via the Darcy–Weisbach equation (Larock, Jeppson, & Watters, 1999) from the pressure drop measured at a known length of the pipeline under nominal operating conditions, which is equivalent to considering Eq. (1) under steady-state conditions. Subsequently, the coefficient of relative roughness ϵ of the pipeline was calculated from the prior value of the friction by solving the Colebrook–White equation (Genić et al., 2011). The current value of the friction in this work is calculated by using the Swamee–Jain equation from the relative roughness calculated from the prior friction factor.

The wave velocity c was estimated from the elastic properties of the fluid and the pipe. The estimated values of the flow parameters in the pipeline, as well as other properties and the experimental conditions, are summarized in Table 1. The complete procedure for estimating these values is described in Bermúdez et al. (2017).

In order to transform (1)–(2) into a simpler system of ordinary differential equations, both equations are discretized in the spatial variable z . Henrie et al. (2016) reports that the most used discretization

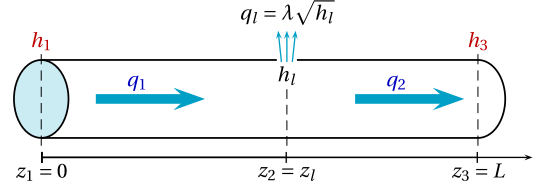


Fig. 4. Variables in the dynamic model of pipeline.

method, in model-based methods for real-time applications, is the finite difference ($\partial q / \partial z \approx \Delta q / \Delta z$) with a constant step size Δz . In contrast, the method proposed here considers first-order finite differences with variable step size, running through the complete pipeline in only two steps, because the real-time execution needs to keep the computational effort as low as possible. In this way, by dividing the continuous interval $z \in [0, L]$ into a three-point discrete partition $\{z_k\} := \{0, z_l, L\}$, where z_l represents the unknown leak position, and considering a leakage flow rate $q_l = \lambda \sqrt{h(z_l, t)}$, the dynamic model of the pipeline is expressed by the ODE system:

$$\dot{q}_1 = \frac{gA}{z_l} (h_1 - h_2) - \frac{f(q_1)}{2Ad} q_1 |q_1|, \quad (3a)$$

$$\dot{h}_2 = \frac{c^2}{gA z_l} \left(q_1 - q_2 - \lambda \sqrt{|h_2|} \right), \quad (3b)$$

$$\dot{q}_2 = \frac{gA}{L - z_l} (h_2 - h_3) - \frac{f(q_2)}{2Ad} q_2 |q_2|, \quad (3c)$$

where $q_1 \stackrel{\text{def}}{=} q(0, t)$, $q_2 \stackrel{\text{def}}{=} q(L, t)$, $h_1 \stackrel{\text{def}}{=} h(0, t)$, $h_2 \stackrel{\text{def}}{=} h(z_l, t)$ and $h_3 \stackrel{\text{def}}{=} h(L, t)$. In the remainder of the paper, the alias h_l (“ h at leak position”) will also be used to refer to h_2 . The inlet pressure h_1 and the outlet pressure h_3 are assumed to be known externally determined by the pump power, but the pressure h_2 at the leak point and the inlet and outlet flow rates (q_1 and q_2) are unknown dynamic variables. The model for q_l is taken from the Torricelli equation for hole leakage (Verde et al., 2013), and the absolute value in the square root of (3b) is included to avoid domain errors when operating numerically with it. The coefficient λ depends on the diameter of the hole where the leak occurs and allows us to size it.

Although theoretically z_l can be located at any point between the ends of the pipeline, in practice it is initially assumed at the midpoint, but can be dynamically adjusted to match the actual position of the leak. Fig. 4 shows the arrangement of the variables that constitute the dynamic model, where h_1 and h_3 are assumed to be known inputs.

2.2. Leak diagnosis by using the EKF

The basic component for the proposed leak diagnosis system is an Extended Kalman Filter (EKF), which allows estimation of leakage parameters (λ and z_l) and filtering variations of measurement noise in the pressure and flow signals. Basically, the Kalman Filter (Kalman, 1960) is a state estimator that allows optimum calculation of unmeasured states

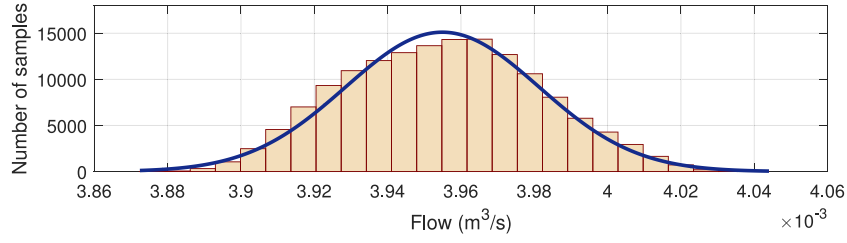


Fig. 5. Gaussian distribution of measurement noise.

in a linear dynamic system from a statistical characterization of the noise at measured outputs. The EKF (Cox, 1964) is a generalization of Kalman Filter for nonlinear systems presenting small or moderate nonlinearities, as in the pipeline model.

The following stochastic linearized model is considered to apply the EKF in a noisy nonlinear system with discrete-time deterministic dynamic component $\mathbf{x}_{k+1} = \Phi(\mathbf{x}_k, \mathbf{u}_k)$, for state transitions, and static component $\mathbf{y}_k = \eta(\mathbf{x}_k)$, for measurements:

$$\mathbf{x}_{k+1} = \mathbf{A}_k \mathbf{x}_k + \mathbf{B}_k \mathbf{u}_k + \mathbf{w}_k, \quad (4)$$

$$\mathbf{y}_k = \mathbf{C}_k \mathbf{x}_k + \mathbf{v}_k, \quad (5)$$

where \mathbf{x} are the states (variables that determine the system dynamics), \mathbf{y} are the noisy output measurements, \mathbf{w} is the process noise, and \mathbf{v} is the measurement noise. It is assumed that both types of noise are zero-mean and white Gaussian. The coefficients \mathbf{A} and \mathbf{B} are the Jacobian matrices obtained from the derivatives of the transition function Φ , and \mathbf{C} is the observation matrix related to the measurement function η .

In the experimental pipeline, spectral analysis determined that the frequency components of the noise in the pressure and flow signals are not limited to a specific frequency range but rather are distributed almost evenly over the Nyquist interval $[0, \frac{1}{2}f_s]$, whose upper bound depends on the sampling rate f_s . In addition, the statistical distribution of the measurements at the nominal operating conditions of the pipeline showed a nearly normal “flared” shape, as shown in Fig. 5 for the flow rate signal. This features (whiteness and Gaussianity) of the noise and the nonlinearity of the system (3a)–(3c) justifies the use of an EKF as a state observer to estimate the leak parameters in the pipeline.

To do that, those parameters (z_i, λ) are considered to be new state variables with dynamics: $\dot{z}_i = 0, \dot{\lambda} = 0$, and which can be included in the original state vector. Thus, the new augmented state vector is: $\mathbf{x} = [q_1 \ h_2 \ q_2 \ z_1 \ \lambda]^T = [x_1 \ x_2 \ x_3 \ x_4 \ x_5]^T$.

In this way, the extended dynamic model of the pipeline-leak assembly is formed by

$$\dot{x}_1 = \frac{gA}{x_4} (u_1 - x_2) - \frac{f(x_1)}{2Ad} x_1 |x_1|, \quad (6a)$$

$$\dot{x}_2 = \frac{c^2}{gA x_4} (x_1 - x_3 - x_5 \sqrt{|x_2|}), \quad (6b)$$

$$\dot{x}_3 = \frac{gA}{L - x_4} (x_2 - u_2) - \frac{f(x_3)}{2Ad} x_3 |x_3|, \quad (6c)$$

$$\dot{x}_4 = 0, \quad (6d)$$

$$\dot{x}_5 = 0, \quad (6e)$$

where the pressures at the ends of the pipeline are taken as control variables of the dynamic model, so that the input vector is $\mathbf{u} = [h_1 \ h_3]^T$.

The conversion of (6a)–(6e) to an explicit discrete-time model suitable to apply EKF was obtained by applying the improved Euler’s method (also referred as Heun’s method) to this continuous-time model that has the form $\dot{\mathbf{x}} = \phi(\mathbf{x}, \mathbf{u})$:

$$\mathbf{x}_{k+1} = \Phi(\mathbf{x}_k, \mathbf{u}_k) = \mathbf{x}_k + \frac{T_s}{2} (\phi(\mathbf{x}_k, \mathbf{u}_k) + \phi(\mathbf{x}_k + T_s \phi(\mathbf{x}_k, \mathbf{u}_k), \mathbf{u}_k)) \quad (7)$$

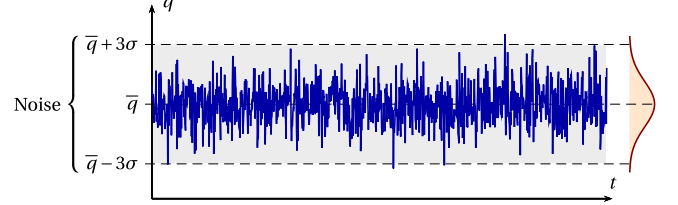


Fig. 6. Measurement noise characterization.

The model required to apply the EKF is completed with the observation Eq. (8) that relates the state \mathbf{x} to the measurement $\mathbf{y} = [q_1 \ q_2]^T$:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}}_{\mathbf{C}} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \quad (8)$$

The uncertainty in the measurements is expressed by a diagonal covariance matrix \mathbf{R} , whose components were determined by calculating the variance in the values measured by each sensor during two minutes for the nominal operating conditions (in steady state), in which it is assumed that the arithmetic mean of the measurements represents the true value and the variation around it is the noise, as shown in Fig. 6. Thus

$$\mathbf{R} = \text{diag}([\sigma_{q_1}^2, \sigma_{q_2}^2]) \quad (9)$$

The EKF estimator has been implemented in two MATLAB sub-routines: TimeUpdate() that performs an a priori estimate (rough) predicting the state value for the next time interval, and MeasurementUpdate() that performs a posteriori (fine) estimate by incorporating the most recent measurements. The implementation of the EKF algorithm is as follows:

The Jacobian matrices in (15) are calculated numerically using finite differences in each time step. The friction factor f is also updated in each EKF iteration for both flow rates, q_1 and q_2 , to consider changes in the flow regime due to leakage. Because the determination of f from the implicit Colebrook–White formula (White, 2016) requires an inner iteration, with additional computational effort, instead the explicit approximation of Swamee–Jain Swamee and Jain (1976) is used, since the simplification error is less than 0.2% under nominal operating conditions:

$$f(q) = 0.25 \left(\log_{10} \left(\frac{\epsilon}{3.7} + 5.74 \left(\frac{Av}{qd} \right)^{0.9} \right) \right)^{-2} \quad (21)$$

In practice, the residues adhere to the pipe walls reducing the diameter d and modifying the roughness ϵ . However, the removal of these residues is considered in the maintenance plan of the system. In addition, the proposed methodology could eventually be updated by some identification algorithm to dynamically adjust the roughness and diameter.

Algorithm 1: EKF estimator for the leak parameters

1. Initialize the EKF with the starting values of state, $\hat{\mathbf{x}}_0^-$, and state estimation error covariance, \mathbf{P}_0^- :

$$\hat{\mathbf{x}}_0^- = E(\mathbf{x}_0) \quad (10)$$

$$\mathbf{P}_0^- = E(\mathbf{x}_0 - \hat{\mathbf{x}}_0^-)(\mathbf{x}_0 - \hat{\mathbf{x}}_0^-)^T \quad (11)$$

where $\hat{\mathbf{x}}_k^-$ is the estimated state before incorporating the measurements at the k -th sampling time. Thus, $\hat{\mathbf{x}}_0^-$ is an arbitrary state estimate before any measurement is made. In practice, $\hat{\mathbf{x}}_0^-$ was taken from the nominal operating conditions of the pipeline:

$$\hat{\mathbf{x}}_0^- = [q_{\text{nom}}, (h_{1,\text{nom}} + h_{3,\text{nom}})/2, q_{\text{nom}}, L/2, 0]^T,$$

where the subscript nom indicates values in nominal operating conditions without leakage.

2. For $k = 0, 1, 2, \dots$ (all discrete-time steps during operation), perform the following calculations:

- 2.1. Subroutine MeasurementUpdate: Calculate the Kalman gain \mathbf{K}_k and update the state and state estimation error covariance using the measured data, observation matrix \mathbf{C} defined in (8) and previous estimates $\hat{\mathbf{x}}_k^-$ and \mathbf{P}_k^- :

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{C}^T (\mathbf{C} \mathbf{P}_k^- \mathbf{C}^T + \mathbf{R})^{-1} \quad (12)$$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k (\mathbf{y}_k - \mathbf{C} \hat{\mathbf{x}}_k^-) \quad (13)$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{C}) \mathbf{P}_k^- \quad (14)$$

- 2.2. Subroutine TimeUpdate: Calculate the Jacobian matrix \mathbf{A}_k of the state transition function Φ to update state estimation error covariance and project the state forward in time:

$$\mathbf{A}_k = \left[\frac{\partial \Phi}{\partial \mathbf{x}} \right]_{\mathbf{x}=\hat{\mathbf{x}}_k} \quad (15)$$

$$\hat{\mathbf{x}}_{k+1}^- = \Phi(\hat{\mathbf{x}}_k, \mathbf{u}_k) \quad (16)$$

$$\mathbf{P}_{k+1}^- = \mathbf{A}_k \mathbf{P}_k \mathbf{A}_k^T + \mathbf{Q} \quad (17)$$

In each discrete-time iteration, the diagnosis of the flow conditions in the pipeline is represented by the following three variables:

$$\hat{q}_l := \hat{x}_k[5] \sqrt{\hat{x}_k[2]} \quad \text{Leak flow rate, double precision real-type} \quad (18)$$

$$p_l := \hat{q}_l \geq q_{\text{nom}}/100 \quad \text{Flag for presence of leaks, boolean type} \quad (19)$$

$$\hat{z}_l := \hat{x}_k[4] \quad \text{Leak location, double precision real-type.} \quad (20)$$

Regarding the condition (19), to avoid false alarms caused by instant values of \hat{q}_l that may exceed the threshold $q_{\text{nom}}/100$ due to numerical transients in the dynamic model runs, the estimated leakage flow rate used by the decision logic that discriminates leakage considers a moving average of the estimates over 1 s time interval. For an integer sample rate f_s [Hz], the estimated leakage flow in an arbitrary k th discrete time step is given by

$$\hat{q}_l^{(k)} = \frac{1}{f_s} \sum_{i=0}^{f_s-1} \lambda^{(k-i)} \sqrt{h_l^{(k-i)}}. \quad (22)$$

2.3. Improvements to EKF-based LDI by using a steady-state estimation

In the application of the EKF to detect real leaks in water pipelines, Delgado-Aguinaga and Begovich (2017) have reported some inaccuracies in the estimates of the leak position with a discrepancy of more than 100 m in some cases. To reduce these inaccuracies, the proposed method refines the estimation output by considering a steady-state solution of the dynamic model, which is incorporated when the variables estimated by the EKF tends to reach stationary states.

When the estimated leakage \hat{q}_l is maintained for a considerable time (user configurable, in the diagnosis software), the final estimate of the flow and position leakage is refined from the steady-state model of the pipeline:

$$\frac{H_l - H_1}{z_l} + \frac{f(Q_1)Q_1|Q_1|}{2A^2dg} = 0, \quad (23a)$$

$$\frac{c^2}{gAz_l} (Q_1 - Q_2 - \lambda\sqrt{H_l}) = 0, \quad (23b)$$

$$\frac{H_3 - H_l}{L - z_l} + \frac{f(Q_2)Q_2|Q_2|}{2A^2dg} = 0, \quad (23c)$$

where the capital letters Q_k and H_k indicate steady-state values of the dynamic variables q_k and h_k . Eqs. (23a)–(23c) were obtained from the spatially discretized Eqs. (3a)–(3c) by replacing with zero the time derivatives and using $\Delta q/\Delta z$ to approximate the spatial partial derivatives $\partial q/\partial z$. The solution of this nonlinear system is obtained using the iterative Levenberg–Marquardt (LM) algorithm (Moré, 1978).

Given the need to ensure a rapid convergence of the LM procedure, the iterative process is initialized with the approximate solution $(z_l^{(0)}, H_l^{(0)}, \lambda^{(0)})$ obtained analytically from steady-state equations (23a)–(23b) assuming a non-leak nominal friction factor f_{nom} :

$$z_l^{(0)} = \frac{2A^2dg(H_1 - H_3) - f_{\text{nom}}LQ_2^2}{(Q_1^2 - Q_2^2)f_{\text{nom}}} \quad (24a)$$

$$H_l^{(0)} = \frac{f_{\text{nom}}LQ_1^2Q_2^2 + 2A^2dg(H_3Q_1^2 - H_1Q_2^2)}{2A^2dg(Q_1^2 - Q_2^2)} \quad (24b)$$

$$\lambda^{(0)} = \frac{Q_1 - Q_2}{\sqrt{H_l^{(0)}}} \quad (24c)$$

The steady-state solution $(z_l, q_l = \lambda\sqrt{H_l})$ obtained from (23a)–(23c), using the LM algorithm, is merged with the dynamic state estimation obtained from the EKF observer using a Bayesian data fusion (Klein, 2012). The data fusion procedure consists in taking a

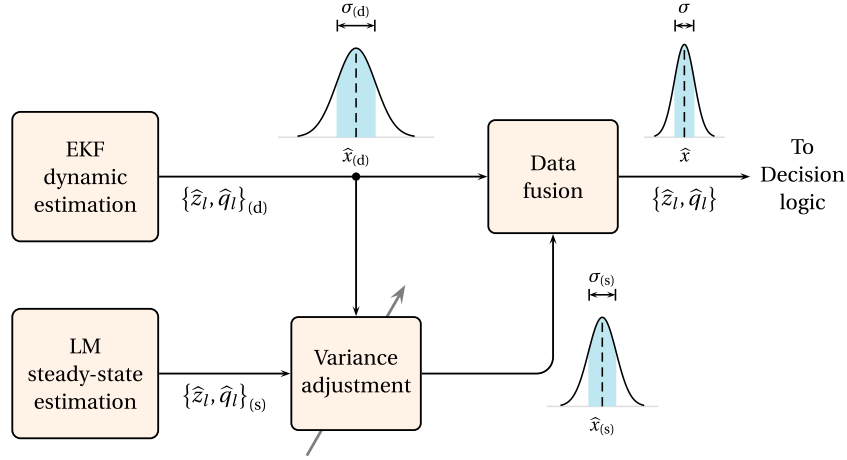


Fig. 7. Data fusion scheme.

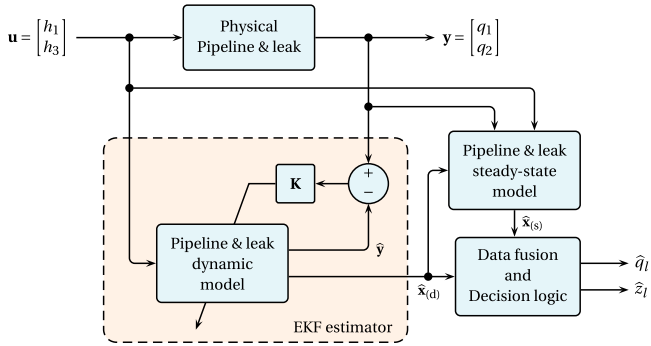


Fig. 8. General scheme of the LDI system.

moving weighted average of both estimates (steady-state and dynamic), using the variances of both estimates as weighting factors. The variance of the dynamic estimates, $\sigma_{(d)}^2$, is continuously updated within the EKF algorithm; but the variance of the steady-state estimates, $\sigma_{(s)}^2$, is approximated empirically in each discrete time-step from the rate of change of leak variables. For this, a sigmoid function is used to decreases $\sigma_{(s)}^2$ from $\sigma_{(s),\max}^2$ to $\sigma_{(s),\min}^2$ as the leakage dynamics passes from the transient behavior to the steady state. The minimum values of $\sigma_{(s)}^2$ were obtained experimentally from the uncertainty in the results of the steady-state solution of a series of tests with stationary leaks. The maximum value of $\sigma_{(s)}^2$ was arbitrarily set as $\sigma_{(s),\max}^2 := 10000 \sigma_{(s),\min}^2$. The formula used to combine the steady-state and dynamic estimates, $\hat{x}_{(d)}$ and $\hat{x}_{(s)}$, generating a single estimate \hat{x} , is

$$\hat{x} = \frac{\sigma_{(d)}^2 \hat{x}_{(s)} + \sigma_{(s)}^2 \hat{x}_{(d)}}{\sigma_{(d)}^2 + \sigma_{(s)}^2}, \quad (25)$$

where $x \in \{z_l, q_l\}$ indicates a decision variable for leak diagnosis. The variance of the combined estimate is given by [Raol \(2016\)](#):

$$\sigma^2 = \frac{\sigma_{(d)}^2 \sigma_{(s)}^2}{\sigma_{(d)}^2 + \sigma_{(s)}^2}. \quad (26)$$

From a statistical viewpoint, and assuming randomness in steady-state and dynamic estimation errors, Eq. (26) shows that the uncertainty after combining both estimates is always lower than the individual uncertainty of each one, which guarantees a final estimation more accurate or, at least, less noisy. The data flow for the data fusion is illustrated graphically in Fig. 7. The representation of a more slender Gaussian in the output of merged data corresponds the fact that the uncertainty in the combined estimate has been reduced.

Fig. 8 graphically shows the interactions between the physical pipeline, its dynamic model operated by the EKF and its steady-state model, within the diagnosis system.

The complete leak diagnosis procedure, as programmed in the software, is summarized in Algorithm 2.

Algorithm 2: General leak-diagnosis procedure

For Each discrete-time Do

1. Read the pressure and flow sensors/transmitters through the DAQ system.
2. Compute the dynamic state estimation with EKF, as described in Algorithm 1, and compute the steady-state estimation by solving (23a)–(23c) using the LM algorithm.
3. From the dynamic estimation, compute $\hat{q}_l = \hat{\lambda} \sqrt{\hat{h}_l}$ and the actual change rate of \hat{q}_l . Then, determine if the regime is steady-state or transitional.
 - If the regime is steady-state setting up the maximum variance of steady-state estimates, Else adjust the variance of the steady-state estimates according to the change rate of \hat{q}_l .
4. Compute the combined estimate for \hat{z}_l and \hat{q}_l by using data fusion according to (25).
5. Decision logic for diagnosis purposes:
 - If $\hat{q}_{\text{leak}} \geq \frac{1}{100} q_{\text{nom}}$, trigger an alarm and show leak information: \hat{z}_l and \hat{q}_l . The user can change the threshold to adjust the system sensitivity.
6. Save measurement data, steady-state and dynamic estimates, and diagnostics, for offline visualization or analysis.

End For

3. Experimental results and discussion

Since the effectiveness of EKF depends fundamentally on the process model and a correct statistical characterization of the noise, the validity of the dynamic model was verified before the leak diagnosis system was tested using the RMS error to quantify the discrepancy between

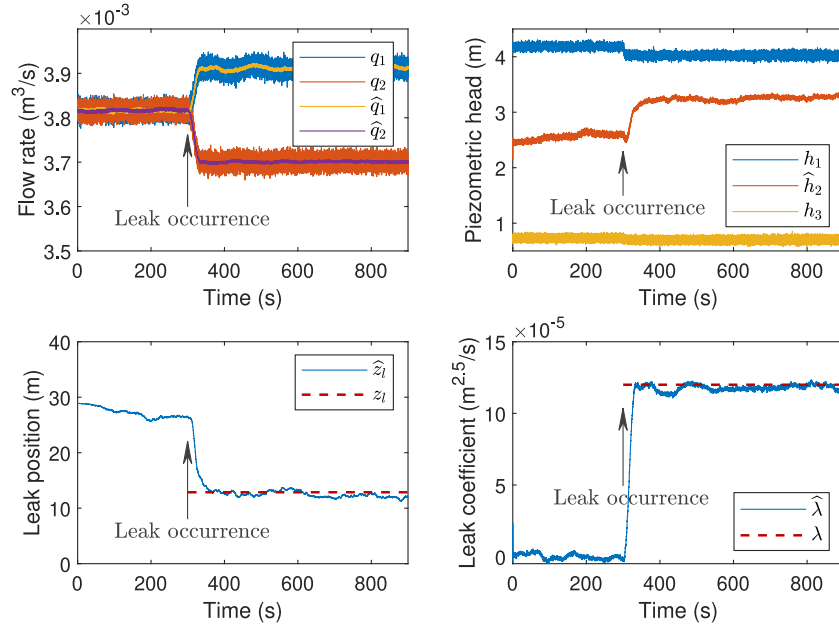


Fig. 9. Convergence of the EKF when locating a leak at position $z_2 = 12.87$ m, induced at time $t = 300$ s.

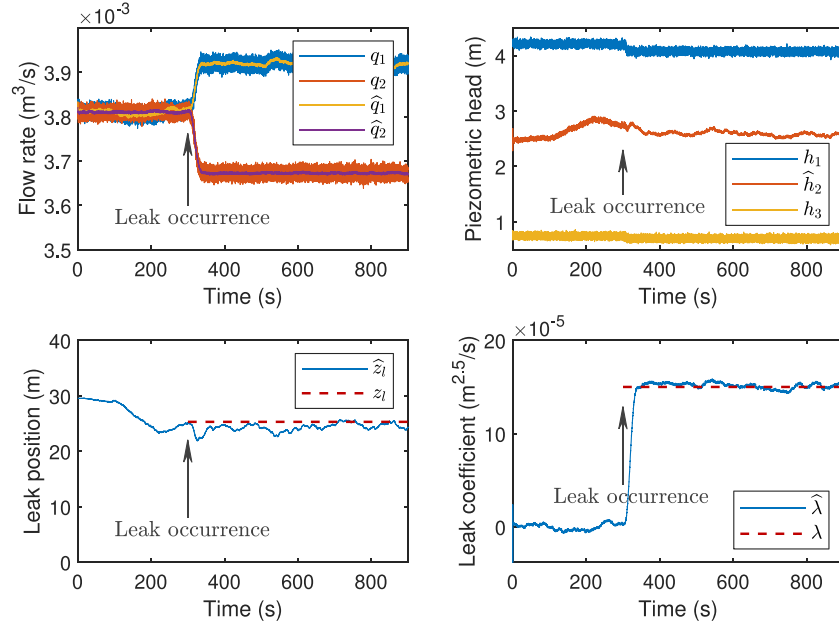


Fig. 10. Convergence of the EKF when locating a leak at position $z_3 = 25.30$ m, induced at time $t = 300$ s.

the experimental measurements and the dynamic model of the pipeline-leak system. Considering the estimated pipeline parameters shown in Table 1, the error in the model (3a)–(3c) with respect to the measurements in the physical plant resulted in $e_{\text{RMS}} = 0.24\%$ for steady-state without leakage, and $e_{\text{RMS}} = 0.41\%$ during transients caused by leakage.

The complete LDI algorithm was initially tested at two different sampling rates, 100 Hz and 1000 Hz. The main difference in the results of these two cases is that at 100 Hz the EKF estimator presents some convergence problems at the pipeline ends. These problems become more important as the sampling rate decreases, making it difficult to locate the leak and it causes both false positives and false negatives. The correctly detected leakage rate was 100% for leaks at positions 2, 3 and 4 of the experimental pipeline. For position 1 only 75% were correctly detected at 1000 Hz, none in this position was detected by EKF at 100 Hz. The state estimates with the EKF were obtained with the

noise variances $\mathbf{Q} = \text{diag}([1 \times 10^{-5}, 1 \times 10^{-2}, 1 \times 10^{-5}, 2500, 1 \times 10^{-6}])$ and $\mathbf{R} = \text{diag}([5 \times 10^{-3}, 5 \times 10^{-3}])$. Figs. 9 and 10 show the convergence of the EKF in the estimation of the leakage parameters for induced leaks in valves L_2 and L_3 of the experimental pipeline. As described in Algorithm 2, the variation observed in the estimated values \hat{z}_l and $\hat{\lambda}$ is smoothed by a data fusion procedure with the steady-state solution when they are displayed in the user interface of the diagnosis program. Fig. 11 shows a screenshot of a typical leak diagnosis session, and Fig. 12 graphically shows the observer performance on the leak location, before and after incorporating the steady-state estimate. The smoothing variation on \hat{z}_l incorporating the steady state solution corresponds to a reduction near of 78% compared with the EKF approach. In other words, only the 22% of the variation remains after the data fusion. This residual variation can be attributed to measurement noise and unknown dynamic components of the process that were not modeled. Regarding

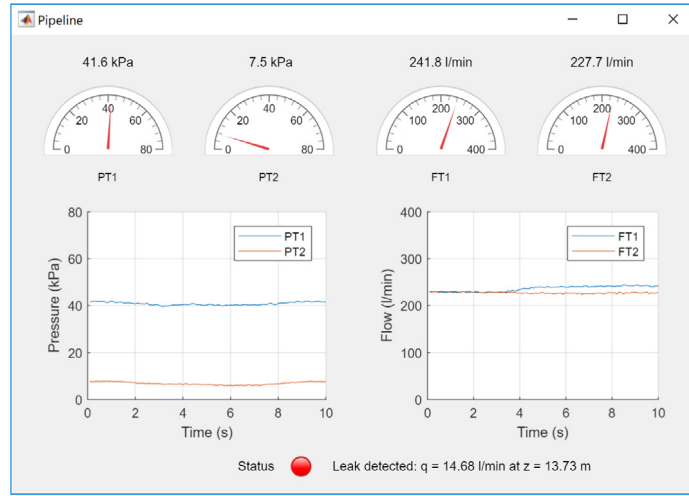


Fig. 11. Snapshot of a diagnosis session.

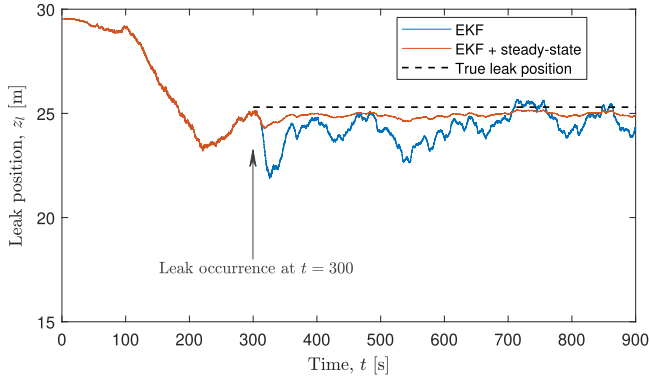


Fig. 12. Improvements in leak location when incorporating the steady-state solution.

Table 2
Summary of leak location results for $f_s = 1000$ Hz.

Real position	Number of tests (*)	Average estimated position	Position error
$z_1 = 2.36$ m	4	7.25 m	8.5%
$z_2 = 12.87$ m	4	13.59 m	1.3%
$z_3 = 25.30$ m	4	26.34 m	1.8%
$z_4 = 41.14$ m	4	42.68 m	2.7%

(*) The same number of runs were also performed without leaks, and no false alarms were presented.

the estimation of \hat{z}_l , the accuracy was improved by 0.69% with respect to the use of the EKF alone, which represents an important improvement with respect to the reported method in Delgado-Aguinaga et al. (2015).

Table 2 summarizes the results on the accuracy of the leak location; errors are presented as percentages of pipeline length. The error in the leak location increases near the ends of the pipe, so the errors reported in Table 2 are large for leaks that were intentionally induced near the ends. In all successful cases, the error in the calculated leakage flow rate did not exceed 1% of the nominal flow rate in the pipeline. All tests were performed under turbulent flow conditions, $22\,000 \leq Re \leq 115\,000$. A centrifugal pump was used to set the Reynolds number in that interval.

It is important to mention that a loss of accuracy, on the leak location, was observed near to the pipeline inlet ($z_l \rightarrow 0$) when the sampling rate was decreased from 1000 Hz to 100 Hz. A larger set of experiments were performed for frequencies lower than 100 Hz. The results for a set of 96 experiments are summarized in Table 3. As

Table 3

Effectiveness in the leak location for different sampling rates, within a 3% of maximum error.

Sampling rate $f_s = 1/\Delta t$	Leak positions			
	$z_1 = 2.36$ m	$z_2 = 12.87$ m	$z_3 = 25.30$ m	$z_4 = 41.14$ m
1000 Hz	75%	100%	100%	100%
500 Hz	0%	100%	100%	100%
100 Hz	0%	100%	100%	100%
50 Hz	0%	100%	100%	100%
10 Hz	0%	0%	75%	50%
5 Hz	0%	0%	0%	0%

can be observed, when the sampling rate decreases the percentage of effectiveness decreases as the leakage location approaches either ends of the pipeline. For sampling rates less than 5 Hz, no results are shown because the EKF does not converge due to the short length of the experimental pipeline. However, the EKF has already been applied earlier with lower sampling rates in a real long pipeline, as reported in Delgado-Aguinaga and Begovich (2017). Table 3 only shows the results obtained with the experimental prototype because it was possible to arbitrarily vary both the leak location and the sampling rate.

From the results summarized in the contingency Table 3, it can be observed that there exist a relationship between the maximum time step Δt and the minimum length step Δz affecting the leak location. This link was calculated taking a weighted average on the sampling rates that mark the effective leak-detection boundary (**bold font** in Table 3), using the percentages of effectiveness as weights w_i , which can be expressed by:

$$\frac{\Delta z}{\Delta t} \Big|_{\text{boundary}} = \frac{\sum_{\text{boundary}} w_i \frac{\Delta z}{\Delta t}}{\sum_{\text{boundary}} w_i} = 936.32, \quad (27)$$

where Δt represents the time-step used on sampling and Δz is the distance from the leak position to the nearest pipeline end according to spatial discretization described in Section 2.1. Given that the spatial discretization of the pipeline contains two values of Δz , which are z_l and $L - z_l$, the leak detection condition (27) is reformulated, in terms of the sampling rate f_s and the leak position z_l as:

$$\min(z_l, L - z_l) f_s \geq 936.32. \quad (28)$$

The constraint (28) that imposes a lower limit on the sampling rate (upper limit on the sampling period) to locate leaks at a certain distance, which can be explained due to the discretization error inherent to all finite difference schemes such as those used in Section 2.1 to discretize

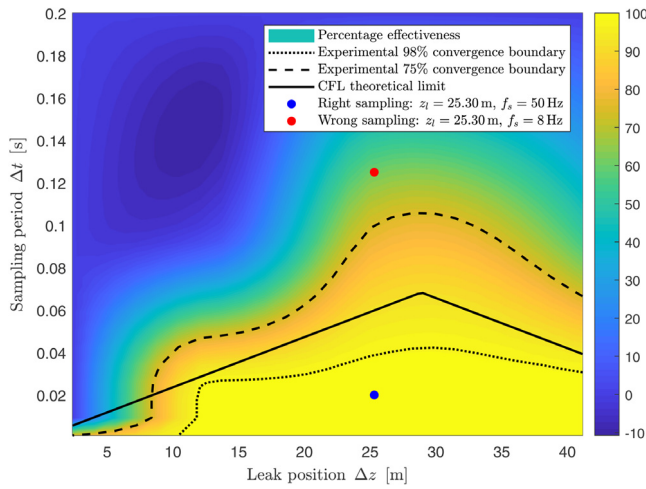


Fig. 13. Convergence region of leak estimator. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

in space and time the partial differential equations that model the flow on which the EKF estimator is based. In particular, for hyperbolic PDEs such as (1)–(2), it has been determined that a condition for the convergence of explicit discretization methods is that the spatial and temporal discretization steps satisfy the so-called Courant–Friedrich–Lewy (CFL) stability condition (Belytschko & Mullen, 1978; Chaudhry, 2014):

$$\Delta z \geq c \Delta t, \quad (29)$$

which shows that to guarantee the convergence of the numerical solution, the size of the time-step must decrease as the size of the length-step decrease. This dependence between the resolutions of time and length domains is shown graphically in Fig. 13. The continuous curve represents the theoretical limit imposed by the CFL condition, for different combinations of Δz and Δt in the experimental pipeline. The percentage of effectiveness in the leak location of the proposed method has been superimposed as a color map. Because the experimental grid consists of only 24 points (those corresponding to Table 3), the remaining values on the grid were obtained by interpolation. Additionally, the level curves corresponding to effectiveness of 75% and 98% have been plotted, along with two isolated points that correspond to typical leakage scenarios, one where the implementation of the proposed method managed to locate the leakage and the other one where it was not possible because the combination leakage-distance/sampling-period was outside the limits of convergence. For example, considering a leak in the valve L_2 at $z = 25.30$ m, the estimated effectiveness decreases from 100% (blue dot in Fig. 13) to less than 75% (red dot) when the sampling rate changes from 50 Hz to 8 Hz.

4. Conclusion and remarks

An online EKF-based leak detection and location system for pipelines has been described in this paper. The dynamics of the transported fluid has been modeled in a spatial partition of only three points along the pipe, with the position of the leak at the variable midpoint. The application of EKF produced good results in an experimental pipeline. In general, the EKF estimator converges correctly to the flow rate of the leak, although it was determined that this methodology presents problems of convergence to the position of the leak near the ends of the pipeline. This deficiency is explained by the dependence between temporal and spatial step sizes, as established by the CFL condition. Theoretically, the convergence of the LDI system near the pipeline ends

can be improved by decreasing the sampling period, but in practice limitations were found because the real-time implementation requires a processing time between each time step and also because most SCADA systems in real water distribution systems do not operate at very high sampling rates. In order to improve the leak detectability near the pipeline ends, without the requirement to decrease the sampling period, an alternative that can be explored in future work is the reformulation of the EKF algorithm to use implicit discretization methods or a finite difference scheme of higher order.

Incorporating the steady-state solution into the diagnosis procedure allows a more accurate leak location than when only the EKF output is used. Also, initialization of the steady-state calculations with the analytical solution under nominal conditions allowed rapid convergence within the sampling period required for real-time implementation. The improvement in the accuracy achieved by combining the EKF output with the steady-state estimate may seem, at first glance, insignificant (less than 1%, less than one meter in the prototype pipeline). However, if a full-scale pipeline is considered, this small percentage improvement would result in several meters, which is equivalent to considerable monetary and time savings when locating and repairing the leak in the case of underground pipelines.

Acknowledgment

This research has been supported by the Consejo Nacional de Ciencia y Tecnología (CONACyT), Mexico under the grant PN-2016/3595.

References

- API, (2002). Api 1130: Computational pipeline monitoring for liquid pipelines.
- Arifin, B., Li, Z., Shah, S. L., Meyer, G. A., & Colin, A. (2018). A novel data-driven leak detection and localization algorithm using the kantorovich distance. *Computers & Chemical Engineering*, 108, 300–313.
- Belytschko, T., & Mullen, R. (1978). Stability of explicit-implicit mesh partitions in time integration. *International Journal for Numerical Methods in Engineering*, 12, 1575–1586.
- Bermúdez, J. R., Santos-Ruiz, I., López-Estrada, F. R., Torres, L., & Puig, V. (2017). Diseño y modelado dinámico de una planta piloto para detección de fugas hidráulicas. In *Congreso nacional de control automático* (Vol. 1) (pp. 2–7). Asociación Mexicana de Control Automático.
- Billmann, L., & Isermann, R. (1987). Leak detection methods for pipelines. *Automatica*, 23, 381–385.
- Camacho-Navarro, J., Ruiz, M., Perez, O., Villamizar, R., & Mujica, L. (2015). Pipe leaks classification by using a data-driven approach based on features from cross-correlated piezo-vibration signals. *Journal of Physics: Conference Series*, 628, 012128.
- Chaudhry, M. H. (2014). *Applied hydraulic transients* (3rd ed.). Springer-Verlag New York.
- Cox, H. (1964). On the estimation of state variables and parameters for noisy dynamic systems. *IEEE Transactions on Automatic Control*, 9, 5–12.
- Delgado-Aguíñaga, J. A., & Begovich, O. (2017). *Water leak diagnosis in pressurized pipelines: A real case study* (pp. 235–262). Springer International Publishing.
- Delgado-Aguíñaga, J., Besançon, G., & Begovich, O. (2015). Leak isolation based on extended kalman filter in a plastic pipeline under temperature variations with real-data validation. In *23th mediterranean conference on control and automation* (pp. 316–321). IEEE.
- Delgado-Aguíñaga, J., Besançon, G., Begovich, O., & Carvajal, J. (2016). Multi-leak diagnosis in pipelines based on extended kalman filter. *Control Engineering Practice*, 49, 139–148.
- Dulhoste, J. F., Guillén, M., Besançon, G., & Santos, R. (2017). *One-dimensional modeling of pipeline transients* (pp. 63–81). Springer International Publishing.
- Genić, S., Arandelović, I., Kolendić, P., Jarić, M., Budimir, N., & Genić, V. (2011). A review of explicit approximations of Colebrook's equation. *FME Transactions*, 39, 67–71.
- Henrie, M., Carpenter, P., & Nicholas, R. E. (2016). *Pipeline leak detection handbook*. Gulf Professional Publishing.
- Hernández, N. (2017). *El robo de combustible: asalto a la nación*, Technical report. Centro de Estudios Sociales y de Opinión Pública, Cámara de Diputados - México.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82, 35–45.
- Klein, L. A. (2012). *Sensor and data fusion: A tool for information assessment and decision making* (2nd ed.). SPIE Press.
- Larock, B. E., Jeppson, R. W., & Watters, G. Z. (1999). *Hydraulics of pipeline systems* (1st ed.). CRC Press.
- Mashford, J., Silva, D. D., Marney, D., & Burn, S. (2009). An approach to leak detection in pipe networks using analysis of monitored pressure values by support vector machine. In *Third international conference on network and system security, 2009. NSS'09* (pp. 534–539). IEEE.

- Moré, J. J. (1978). The Levenberg-Marquardt algorithm: implementation and theory. In *Numerical analysis* (pp. 105–116). Springer.
- Murway, P.-S., & Silea, I. (2012). A survey on gas leak detection and localization techniques. *Journal of Loss Prevention in the Process Industries*, 25, 966–973.
- OECD, (2016). Water governance in cities. *OECD Stud. Water*.
- Oswald, Ú. (2017). Energy security, availability, and sustainability in Mexico. *Revista Mexicana de Ciencias Políticas y Sociales*, 62, 155–196.
- Raol, J. R. (2016). *Data fusion mathematics: Theory and practice*. CRC Press.
- Soldevila, A., Blesa, J., Tornil-Sin, S., Duviella, E., Fernandez-Canti, R. M., & Puig, V. (2016). Leak localization in water distribution networks using a mixed model-based/data-driven approach. *Control Engineering Practice*, 55, 162–173.
- Soldevila, A., Fernandez-Canti, R. M., Blesa, J., Tornil-Sin, S., & Puig, V. (2017). Leak localization in water distribution networks using bayesian classifiers. *Journal of Process Control*, 55, 1–9.
- Swamee, P. K., & Jain, A. K. (1976). Explicit equations for pipe-flow problems. *Journal of the Hydraulics Division*, 102, 657–664.
- Torres, L., Verde, C., Besançon, G., & González, O. (2014). High-gain observers for leak location in subterranean pipelines of liquefied petroleum gas. *International Journal of Robust and Nonlinear Control*, 24, 1127–1141.
- Verde, C. (2001). Multi-leak detection and isolation in fluid pipelines. *Control Engineering Practice*, 9, 673–682.
- Verde, C. (2004). Minimal order nonlinear observer for leak detection. *Journal of Dynamic Systems, Measurement, and Control*, 126, 467–472.
- Verde, C., Gentil, S., & Morales-Menéndez, R. (2013). *Monitoreo y Diagnóstico Automático de Fallas en Sistemas Dinámicos*. Trillas.
- Verde, C., & Torres, L. (Eds.). (2017). *Applied condition monitoring: Vol. 7. Modeling and monitoring of pipelines and networks: advanced tools for automatic monitoring and supervision of pipelines*. Springer International Publishing.
- White, F. (2016). *Fluid mechanics* (8th ed.). McGraw-Hill Education.